HOWARD B. NEWCOMBE

Biology Branch, Chalk River Nuclear Laboratories, Chalk River, Ontario

The Study of Mutation and Selection in Human Populations*

INTEREST IN THE genetics of human populations may be either of an *historical* or of a *predictive* kind. That is to say, most population genetic studies relate fairly directly to one or other of two sorts of question, namely "How did we get this way?" and "Where are we heading?" Both emphases imply a certain amount of interpretation and speculation, but both have served to stimulate the gathering of relevant empirical data of quite non-speculative kinds. In view of the sponsorship of the present lecture by the Eugenics Society, and the special association with the name of Galton, it will deal primarily with the second of these two questions.

To those investigators who are concerned with current genetic trends and their long-term consequences, the large amount of information about people which is gathered routinely for a variety of purposes in modern societies may serve either as a special spur or, because of the immensity and relative inaccessibility of the resulting accumulation of facts, as a major source of discouragement. The difference depends on the view taken of the respective magnitudes of the difficulties as compared with the opportunities for extracting from the existing documentation genetic facts of the kinds needed for the more important sorts of human population studies.

The opportunities are of course substantial, even if as yet largely unrealized, because for no organism other than man have the facts of procreation and family composition, and of mortality and morbidity, been so minutely documented over such large populations. Furthermore, in no experiment with laboratory mammals is it even remotely conceivable that so much money and effort could ever be spent identifying the natures and causes of deviations from physical and mental well being, and social adjustment, as is done in any advanced human society.

A major part of the resulting documentation is, in fact, relevant to studies of those influences which operate to alter the collective germplasm, or pool of genetic material, and which therefore determine the hereditary basis of health and well being of future generations. However, genetic studies are, virtually by definition, family studies; and the biggest barrier to the use of much of the enormous body of recorded facts about people for any sort of genetic research lies in the difficulty of relating, or linking, these facts to the family histories of the persons concerned.

The present account will deal with the nature of the information needed for studies of current genetic trends, and with the ways that may be used to extract and apply such information. As we proceed, the difficulties will seem to loom larger. And yet, to admit to the possibility that we may never know in any substantial detail what changes are in fact taking place in the gene pool, and what factors are potent in influencing these changes for better or worse, would seem quite improper until we get much closer than we are now to exhausting the information that is already potentially at our disposal.

The Nature of Population Genetics

The kinds of information that are required may best be understood by considering for a moment the changes that take place, over a

* The Galton Lecture, delivered in London on 5th May 1965.

single generation cycle, in the collective germ plasm of which each generation is, in turn, the temporary custodian.

The diversity of hereditary characteristics represented at the outset owes its presence to past events of *mutation*, that is of *de novo* hereditary changes, and also, if one thinks of a population less than that of the world as a whole, of migration into the population of genetically diverse people. Such occurrences, in each generation, contribute further to the genetic diversity that has been already accumulated over the past.

The outward expressions of the resulting heterogeneity within the collective pool of genes are only in part determined by the external environment; just as important are the so-called patterns of mating among the immediate ancestors of the generation under observation. If marriages were to occur in a strictly random fashion between members of various population subgroups, the effect would be a blending of many widely different inheritances, so as to obscure much of the hereditary diversity present in the gene pool. In practice however, matings tend, in varying degrees, to occur preferentially within the various population subgroups, as based on family relationships, religious affiliations, ethnic origins and economic circumstances. There is in addition a measurable tendency for persons of similar physical and mental characteristics to marry one another. The effect of these inbreedings and assortative matings, when they occur, is to provide increased opportunity for outward expression of the genetic diversity inherent in the collected inheritance. An extended range of phenotypes is thus subjected to the tests of survival and procreation.

The composition of the gene pool of the next generation depends, of course, upon the selective effects of the differences in mortality and fertility or, in other words, upon what kinds of people reproduce, and how much. Selection does not operate directly upon the genes, to favour or suppress the propagation of one as compared with an alternative gene. Rather, it is the phenotypes or outward expressions, which are in part determined by the genes but which are also frequently influenced by the environments as well, that are "selected" on the basis of their abilities to survive and reproduce. Genes that show a tendency to be associated with phenotypes of high selective value will, as a result, be transmitted preferentially to the gene pool of the next generation.

From all this it is apparent that human population genetics is concerned with a relatively small number of major phenomena, notably mutation, migration, mating patterns and selection. For the study of any of these, family histories or pedigrees are an essential requirement and the more of such histories that can be marshalled the better.

The point I wish to emphasize at this juncture is that the basic pedigree information is accurately recorded on a continuing basis as part of the vital statistics systems of most countries, and is essentially complete for all who remain within the registration areas. These systems also record our geographic and racial origins, religious affiliations and occupations; and if other routine records such as the census enumeration forms, income tax returns and so on are regarded as further potential sources, the documentation of our social characteristics is more complete still. Similarly, our physical and mental attributes are recorded in even greater detail, partly because of our demands for medical and health services, but in other ways as well, as in the case of records of school performance, enlistment in the armed services, and of unemployment, crime and delinquency.

Such documentation is in most instances tied to the names of the individuals concerned so that, in principle at least, virtually all of it could be integrated with the family pedigrees. Of all the recorded information, however, the most fundamentally important is that which serves to identify immediate relatives, and which is contained in the vital records systems in a form that makes possible the compilation on a statistical scale of histories of marriage, procreation and death in individual families. Thus we already possess much of what is needed for detailed and precise studies of fertility differences, mating patterns and so on, and the scope and reliability of the documentation is increasing continuously.

Access to such information is not always simple. Use of some of the records is restricted

by law, and in most instances it is limited by the sheer magnitude of the files. In addition, there are organizational and technical problems of interrelating, on any substantial scale, the facts contained in two or more independently derived records pertaining to the same person or family.

For such reasons, major advances in the use for research purposes of any substantial fraction of the potentially available information are not to be expected in a hurry; but modest developments are in progress now, and some early products of one of these, relating particularly to mutation and selection will be described.

Pedigrees from Record Linkage

The study from which these products have been derived relates to the Canadian province of British Columbia, with a current population of 1·7 million, in which about 10,000 marriages and 40,000 births occur annually. A ten-year marriage file (1946–55) and a six-year birth file (1952–58) have been used. In addition, there are records of stillbirths over the same six-year period, and of child deaths, together with registrations of handicaps among children born in this period, provided by the exceedingly well run British Columbia Registry of Handicapped Children and Adults.

By methods that have been described fully elsewhere,[4, 5, 9, 10, 11] these registrations have been integrated into the form of family histories, starting in each instance with the marriage record (where this is present in the files), followed by those of any procreations arising out of the marriage, and with the records of handicaps and deaths among offspring of the married couples interpolated behind the birth records of the children to which they relate. Somewhat over 200,000 records are currently in use, and the integrated files will in due course be extended backward to 1946 and forward to include current records as they become available.

The technology is now well developed, and the actual merging and linking operation in which current records are used to update the family groupings on an existing master file can be carried out by an electronic computer exceedingly rapidly and economically.

Mutations

One of the applications of the record linkage technology which I propose to discuss in detail has to do with the possibility of detecting and measuring the consequences of recurrent mutations in the human germ plasm. Most of the genes which arise from these mutational events are believed, with considerable reason, to be harmful; and in recent years much thought has been given to the possible importance of such artificially induced increases in mutation rate as must almost certainly result from exposures of human reproductive tissues to man-made radiation.

The problem has substantial financial overtones, because upper limits to the exposures of whole populations from the future peaceful uses of atomic energy are being suggested now by an authoritative body.[15] There is, as a result, considerable pressure on individual geneticists to assess the likely consequences of the mutations that would arise in human populations from given radiation exposures, even where the amount of the harm cannot be judged to within an order of magnitude or so in either direction.

Such estimates of genetic damage from artificially induced mutations are unlikely to attain increased precision until more is known of the importance of the mutations that occur anyway as a result of natural causes. Some fraction of the current burden of hereditary ill health and handicap, and possibly a substantial fraction of it, is believed to be maintained in the population by repeated mutations of natural origin.

There has been a tendency on the part of geneticists in the past to concentrate on studies of mutation frequencies at particular gene loci which control striking and readily scorable characteristics, such as achondroplastic dwarfism, the heritable eye cancer known as retinoblastoma and the conspicuous nerve tumours of von Recklinghausen's disease or multiple neurofibromatosis, to take a few examples. Although the results of such studies are often quoted, they are essentially disappointing for present purposes, not because of any of the inherent imprecisions, but rather because they relate to a small and quite unrepresentative group of

mutant traits and indicate little indeed concerning the possible importance of mutations throughout the total complement of gene loci.

There is at least one other approach to the detection of the consequences of mutation throughout the genome as a whole. If mutations or mutagenic influences accumulate during the reproductive life spans of individuals, genetically determined harm might be detectably increased in offspring from older parents. This suggestion, which was made originally by Penrose [14] and Sonneborn,[16] was in fact later borne out in the case of one quite striking hereditary disease, namely mongolian idiocy or Down's syndrome, the risk of which is greatly elevated among offspring from mothers who are approaching the end of the reproductive period. The mutation that determines this condition, however, involves not just an alteration of a single gene, but rather a gross chromosomal change consisting usually of the incorporation of an additional chromosome over and above the normal number. Nevertheless, other similar but less striking increases in the risks of various hereditary conditions with advancing maternal or paternal ages have been interpreted as probably due to increased incidences of particular gene mutations in gametes from older mothers and fathers.

I should like at this stage to mention certain difficulties in the use of the approach that may perhaps make the early applications of it seem, at first sight, lighthearted indeed, but which should not be taken to imply that we must forever forgo knowledge of the consequences of mutation, or that the approach should be abandoned in favour of some superior alternative when in fact there may be none.

The first difficulty relates to the availability of suitable empirical data. The early observations were based on quite limited ascertainments of the parental age distributions for children with various genetic disorders, and the control data related to populations that were not necessarily representative of those from which the cases were drawn, and which could have differed substantially in the distributions of parental ages at birth. Such limitations of the data arise from essentially artificial causes, in that very substantial numbers of affected children are known to exist and the necessary parental age information is almost universally recorded in the birth registrations. The difficulty of securing the necessary data arises only because the two sorts of information relating to a particular individual tend to be recorded independently on separate documents of quite different kinds, both of which may at some later date be effectively buried in massive files of accumulated records. As will be seen later, other facts needed for the study of mutation by means of this approach are likewise recorded but tend to be inaccessible for the same reason, and may remain so unless economical means are used to interrelate, on a substantial scale, specific by individual, the information contained in various large, independently derived files.

Our own studies of parental age correlations have been based on registrations of live births, stillbirths, handicaps and deaths among just over 200,000 children born in British Columbia during 1953–58, of whom more than 12,000 were registered either as stillborn, handicapped or as having died by the end of 1959. Record linkage techniques were employed which make use of computers to merge the parental ages and birth orders from the birth records, with the diagnoses from the handicap and death records of the same individuals. Although larger amounts of data were derived in this manner than have been analyzed in previous studies, the results must be regarded as modest in comparison with those that may be obtained through applications of similar methods on a continuing basis to these and other files of vital and health records. Furthermore, the most important objective of the study is judged to be the continuing development and testing of methods by which such information may be more fully utilized in the future.

The results serve, however, to illustrate the kinds of effect on risks to the offspring, that are not causally related to the advancing age of the father or mother, but are only secondarily associated with parental age, and which therefore must be carefully distinguished if the approach is to be used at all to study the consequences of mutation. Some of these secondary associations will tend to obscure the effects of mutations; others will tend to simulate or mimic them.

In the former category (i.e. those which tend to obscure) are the special risks to firstborn offspring and to offspring of very young mothers, which fortunately can be distinguished from one another by suitable statistical methods. A whole group of diseases peculiar to early infancy, and in particular the intracranial and spinal injuries at birth, have about twice the frequency among firstborn as compared with secondborn children, when effects of maternal age differences are removed (Table 1). Similarly, children of very young mothers, age nineteen and under, are about twice as prone to a wide variety of conditions, which include intracranial and spinal injuries at birth and postnatal asphyxia, as are children of mothers in the next five-year age group (Table 1); this is true even when effects of birth order differences are removed (for an example of the method for removing the effect of one variable or the other, see Table 2).

More serious for present purposes are the effects that can mimic those of increased mutation frequencies with advancing parental ages. Not surprisingly, the infective and parasitic diseases (notably tuberculosis and polio) tend to be more common among children of birth orders three and over, the same being true of postnatal asphyxia (Table 1). As a result, these conditions exhibit a parental age effect that is in reality secondary to the birth order effect. An environmental factor is almost certainly involved, since for both sorts of condition the effect of birth order is most striking when the mother is young (for an example, see Table 3). Possibly this is because the pregnancies have been closely spaced, or because of crowding in the home, or perhaps because the mothers who have had many children while still very young tend to be from a socio-economic group in which the risks of such conditions are high.

TABLE 1

Summary of special risks associated with birth order and parental ages. (From B.C. handicap and death records linked with birth records, and from stillbirth records)

Disease category	Code	Total cases	Relative risk	χ^2 (D.F. = 1)	Reference (page 125)
Firstborn children (*vs.* 2*nd*)					
(excluding maternal age effects)					
Certain diseases of early infancy	760–776	58	2·16	10·6	7
Intracranial and spinal injury at birth	760	123*	1·80	9·1	12
Higher birth orders (3*rd and over vs.* 1*st and* 2*nd*)					
(excluding maternal age effects)					
Infective and parasitic diseases	001–138	146	1·76	9·9	7
Strabismus	384	285	1·58	11·7	12
Other c.m. of nervous system and sense organs	753	175*	1·84	12·1	12
Postnatal asphyxia and atelectasis	762	773*	1·70	39·2	12
Children of very young mothers (0–19 *vs.* 20–24)					
(excluding birth orders effects)					
Categories VI–XII	400–716	58	2·44	8·6	7
Intracranial and spinal injury at birth	760	72*	2·19	9·6	12
Postnatal asphyxia and atelectasis	762	413*	1·66	15·4	12
Children of older mothers (35 *and over vs.* 0–34)					
(excluding birth order effects)					
Mongolism	325·4	191	7·68	172·6	12
Cerebral palsy	351	215	1·84	11·6	12
C.m. of circulatory system	754	868*	1·66	29·1	12
Children of older fathers (40–99 *vs.* 0–39)					
(excluding maternal age effects)					
Diseases of the respiratory system	470–527	955*	1·61	17·8	13
Congenital malformations	750–759	1727*	1·28	9·1	13
Other c.m. of nervous systems and sense organs	753	135*	2·07	7·8	13

* Includes cases ascertained through death and/or stillbirth records.

To distinguish between these alternatives, further information about the social circumstances and the reproductive histories of the families would be needed, much of which is already documented in the vital registrations, census enumerations and elsewhere. There is no certainty, of course, that genetic factors play any substantial role in the causation of these particular conditions, but the examples serve nevertheless to illustrate the ways in which biases can arise. Similar interactions appear to be quite common (Table 4).

Of less obvious origins are the increased risks of strabismus, or "squint", and of certain "other" congenital malformations of the nervous system and sense organs among children of higher birth orders (Table 1) which might, if birth orders were not taken into consideration in the analysis, appear as maternal age effects

TABLE 2

Example of special risks to children of older mothers, when effects of birth order differences are removed

Congenital malformations of the circulatory system

(Comparing mothers aged 35 and over with those under 35)

	Cases		Controls			
Birth order	Older mothers *a*	Young mothers *b*	Older mothers A	Young mothers B	Relative risk *a*B/A*b*	χ^2
1	8	236	2,611	59,675	0·77	...
2	24	189	4,381	53,411	1·55	...
3	43	136	6,039	36,239	1·90	...
4	37	71	5,372	18,755	1·82	...
5	25	28	3,533	8,598	2·17	...
6–9	31	333	5,161	7,519	1·42	...
10	7	0	1,432	627		...
Weighted mean relative risk (D.F. = 1)					1·66	29·1*

* $P<10^{-7}$.

Data from Newcombe and Tavendale, 1964.[12]

TABLE 3

Example of an interaction between birth order and mother's age as influencing the risk to the child

Postnatal asphyxia and atelectasis

(Comparing birth orders 4 and over with 1 to 3)

	Cases		Controls			
Mother's age group	Higher birth order *a*	Lower birth order *b*	Higher birth order A	Lower birth order B	Relative risk *a*B/A*b*	χ^2
0–19	3	86	95	18,584	6·82	...
20–24	42	181	4,715	56,495	2·78	...
25–29	63	86	14,084	47,584	2·48	...
30–34	64	83	16,605	26,662	1·24	...
35–39	41	51	11,474	10,643	0·75	...
40–44	17	12	3,748	2,282	1·06	...
45–49	2	0	276	106		...
Weighted mean relative risk (D.F. = 1)					1·70	39·2*
Heterogeneity (D.F. = 5)						39·9†

* $P<10^{-9}$.
† $P<10^{-7}$.

Data from Newcombe and Tavendale, 1964.[12]

TABLE 4

Summary of interactions between birth order and mother's age as influencing the risk to the child

(Chi squares for heterogeneity; for details of derivation see Table 3)

Code	Cause	Groups compared	χ^2	D.F.	P
325	Mental deficiency	mother's age 0–19/20–24	8·2	2	0·02
330–398	Nervous system	mother's age 0–19/20–24	8·6	3	0·04
351	Cerebral palsy	mother's age 0–19/20–24	14·0	3	0·003
353	Epilepsy	birth order >2/1–2	10·8	4	0·03
751, Y38·2	Spina bifida	birth order>2/1–2	15·7	5	0·008
758	C.m. bone and joint	birth order>2/1–2	10·2	4	0·04
762	Postnatal asphyxia	birth order>2/1–2	39·9	5	0·001

Data from Newcombe and Tavendale, 1964.[12]

of kinds that would mimic the consequences of mutations accumulating in the female germ cells. Genetic causes of another sort might conceivably be involved, since maternal/foetal incompatibilities could very well be operating to increase the likelihood of sensitization of mothers of higher parities, thereby increasing the risks to offspring of higher birth orders. Such effects would not, of course, indicate any influence of parental ageing as such.

Of the increased risks shown to be associated primarily with advancing age of mother, that of mongolian idiocy is known already to be mutational in origin, while that of cerebral palsy could perhaps be attributed to the changes in intra-uterine environment that occur in older mothers. Evidence to rule out such an effect of the prenatal environment is not wholly unobtainable, but would require another generation of pedigree information to permit comparison of the risks in the current generation of children for different ages of the maternal grandmother where the mother's age is not a variable. No one has seriously attempted to carry out such multi-generation comparisons on any substantial scale as relating to particular diseases or groups of diseases, although data have been obtained on correlations of the grandmaternal age with the sex ratio, of a kind indicating possible differences in prenatal mortality due to lethal mutation in the sex chromosomes.[2] Extensions of the approach to include mortality and morbidity of children, specific by disease or broad category of disease, would require application of pedigree information of a kind that is already contained in the accumulated vital records of most countries, but on a scale that would heretofore have seemed impracticable in the extreme.

Observed correlations of risks with advancing age of the father at the time of the birth, independent of the mother's age, cannot, of course, be attributed in the above manner to a changing intra-uterine environment, but other difficulties of interpretation present themselves. One such correlation, involving the risk of congenital malformation, and more particularly the group of such conditions known in the International Classification as other congenital malformations of the nervous system and sense organs, might perhaps be genuinely indicative of increased mutant frequencies in gametes from ageing fathers (Table 1). However, a similarly increased risk of diseases of the respiratory system (mainly deaths from pneumonia and bronchitis) which has been observed among offspring from older fathers seems quite unlikely to be mutational in origin (Tables 1 and 5). A relatively simple explanation is in this instance apparent from examination of the data; most of the deaths from respiratory diseases among British Columbia children occur in North American Indians, and Indian fathers tend to be older than is usual for the rest of the population. The correlation disappears when the data are broken down by Indian versus non-Indian parentage.

Findings of this sort emphasize the importance of distinguishing those correlations with paternal or with maternal age that are secondary in nature, and that depend on differences in the

parental age distribution for population subgroups within which the risks may be higher or lower than for the population as a whole. The principal limiting factor in the identification of such non-mutational effects is not the amount of recorded knowledge of social characteristics, which is really quite extensive. Rather, it is the mechanical difficulty of bringing together such recorded information, as relating to particular individuals, and of doing so on a large enough scale to permit multiple breakdown of the data without spreading them too thinly between the resulting pigeon-holes or matrix cells.

To complicate matters still further, the extent of the difference between the age of the father and that of the mother, is itself correlated with the risk of infant death (Table 6), an effect which is presumably likewise associated with population heterogeneities involving simultaneously the ages of the parents and the risks to the children.

This relatively long list of the difficulties that besets attempts to derive information about the importance of mutation from studies of parental age correlations may at first sight seem discouraging in the extreme. And yet, it is doubtful if there are any better approaches, and we are still in the happy position of having potentially available to us a vastly greater body of recorded personal information, of strictly relevant kinds, than anyone has so far attempted to analyze.

TABLE 5

Example of a special risk to children of older fathers, when effects of maternal age differences are removed, which is known to be secondary to racial inhomogeneity in the population

Respiratory diseases

(Comparing fathers 40 years and older with those of 39 and younger)

	Cases		Controls			
Mother's age group	Older fathers *a*	Young fathers *b*	Older fathers A	Young fathers B	Relative risk *a*B/A*b*	χ^2
0–19	1	108	8	2,238	3·67	...
20–24	4	306	94	8,922	1·36	...
25–29	22	239	427	8,988	2·14	...
30–34	37	124	1,203	5,707	1·59	...
35–39	45	42	1,638	1,774	1·28	...
40–44	21	3	819	176	1·65	...
45–99	3	0	70	1		
Weighted mean relative risk (D.F. = 1)					1·61	17·8*

* $P<10^{-4}$.

Data from Newcombe and Tavendale, 1965.[13]

TABLE 6

Effects of parental age differences, per se, *on the risk of infant death*

Difference in 5-year age group of husband and wife		Relative risk*	χ^2 (D.F. = 1)	P
(Wife younger)	−4 to −6	1·31	7·6	0·006
	−3	1·29	19·8	0·000,01
	−2	1·11	10·9	0·001
	−1	1·11	0·2	...
(Same age group)	0	0·93	11·6	0·000,6
(Wife older)	+1	0·91	3·6	...
	+2	0·87	0·8	...
	+3 to +4	1·62	1·3	...

* I.e. the risk for a given difference in parental age group as compared with that for all other age differences. The values are weighted means of those derived separately for each 5-year group of father.

From analysis of Newcombe and Tavendale, 1965.[13]

Selection

Some part of the hereditary causes of disease is, of course, maintained in the population, not by repeated mutation but by selection; that is by differences in fertility and mortality that operate in some circumstances to favour the perpetuation of genes which in other settings are the cause of ill health or handicap. To those investigators who are especially concerned with estimation of the consequences of artificially induced increases in the mutation rate, it is important to make full use of any information which will serve to identify those conditions that are maintained in the population by selective forces, and to distinguish them from those that are maintained by repeated mutations. To others who are concerned with all of the various influences that may affect the future quality of the gene pool, it is important to be able to identify the social and other factors that alter the directions in which the selective forces operate.

Precision in detecting differences of fertility and mortality depends to a major degree on the quality and accuracy of the available pedigree information. The identifications of family relationships contained in the vital registration systems are almost universal in their extent, covering as they do whole populations, and they are substantially more reliable than similar information which has been derived by personal interview or questionnaire. Because of this unique precision of the family information from the vital records, the most interesting early products of genetic studies employing record linkage methods relate to the detection of deviations from normal fertility and mortality displayed by families in which certain diseases and handicaps of possible genetic interest have occurred; or, in other words, to the study of selection pressures as associated with major causes of ill health.

The investigations of this kind that have been carried out using the vital records have been modest in size and have not as yet made full use of electronic data processing methods. They serve two purposes at this stage, to demonstrate that information on selection can be obtained from the vital registration system, and to give some hint of the kinds of empirical findings that are likely to come out of future applications of the approach to larger files and to a wider range of diseases.

The particular selection studies which will be described were prompted by a current belief that the Rh-negative gene is kept in the population by a tendency on the part of mothers of infants who die of erythroblastosis to "compensate" for their losses by having another child sooner than would otherwise be the case. If there is such a tendency, it is altogether likely that mothers also compensate in the same fashion for early infant deaths of other kinds, and that the phenomenon may be quite widespread.

Reproductive performances of mothers were therefore studied following stillbirths, infant deaths from a number of different causes and also following the births of mongoloid idiots who did not die. Comparisons have in each case been made with the reproductive histories of mothers of similar ages and parities who were not so selected (Figure 1).

There are distinct advantages in being able to use the whole of a birth population as a control in a study of this kind. Approximately 170,000 birth records were available on magnetic tape representing the five-year period 1954–58. Each record of a birth in 1954, of which there were about 35,000, was used to initiate searches by computer for other births to the same mother in that year and in each of the subsequent years. This was a major undertaking, even with the help of a computer, and the total of over 170,000 searches (i.e. 35,000 per year over the five-year period) took eighty hours on a now obsolete machine (Burroughs 205). However, with the more recent machines now in use, two to three hours would be sufficient.

Mothers of stillborn children, in fact exhibited reproductive compensation, but only over a brief period. During the first twelve months following the stillbirth, between two and three times as many subsequent births took place to these mothers as would be expected in view of their ages and the numbers of children which they already had.[11] Over the second year their collective reproductive performances were about average, while in the third and fourth years these declined to about 70 and 60 per cent of

normal. Taking the reproductive histories over the whole four-year period mothers of stillborn children tended to be less fertile than other mothers of the same ages and parities.

Similar studies of conditions other than stillbirth have been carried out more recently.

Mothers of children registered as dying of erythroblastosis, for example, have shown no tendency to compensate for their losses, the observed number of later born children being almost exactly the number expected of them (Table 7). This does not necessarily mean that compensation has not taken place at some period in the past, and perhaps even up to the time when the hereditary nature of the condition came to be generally recognized, but there is certainly no evidence from the present study that it is still occurring.

The same may be said of mothers of children dying of haemorrhagic disease of the newborn although, for this condition, knowledge of the cause would be less of a deterrent to further reproduction (Table 8). The observed number of later born children was again almost exactly what would be expected.

Not all of the conditions studied, however, failed in this manner to be associated with reproductive compensation, and for deaths from

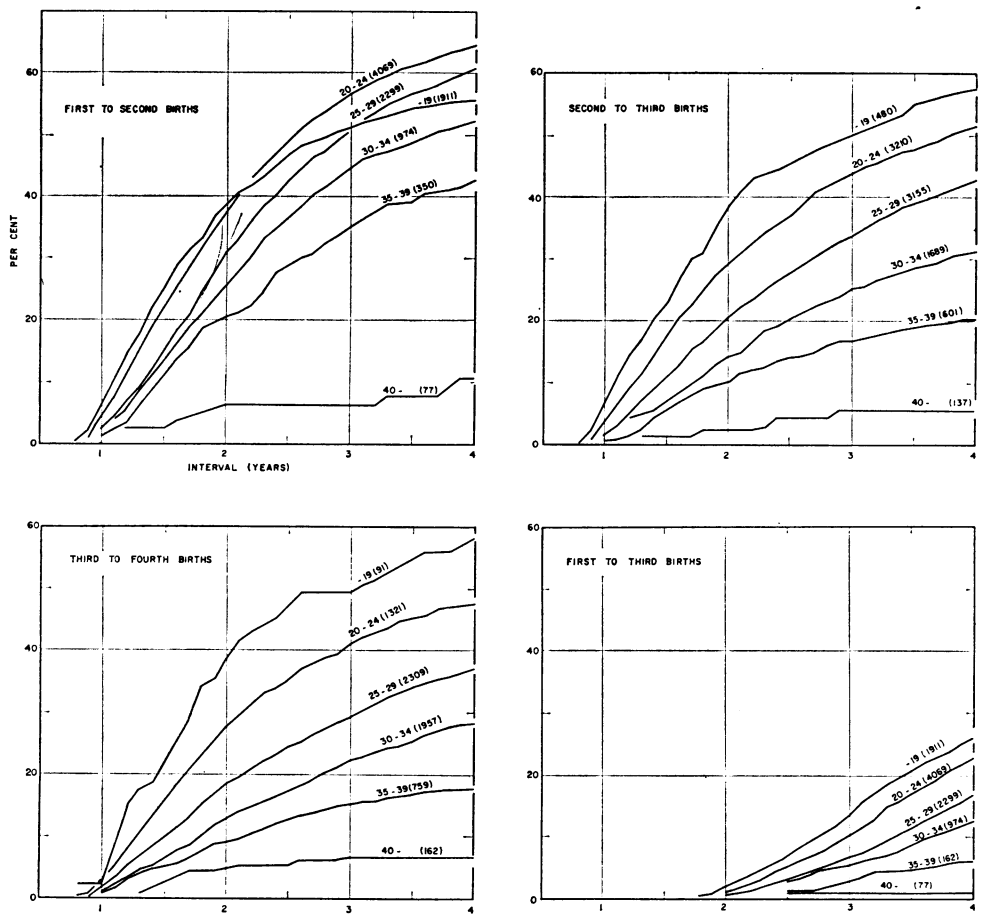

FIGURE 1

Percentage of mothers of liveborn children (born in 1954) having offspring throughout a four-year period following these births.

postnatal asphyxia, which is often associated with failure of the lungs to expand at birth, the effect is striking. From the ninety-seven mothers of affected children, instead of the expected sixty-six later born children, there were 111, an excess of nearly 70 per cent (Table 9).

Consideration of some of the possible causes of such an association serves to illustrate the complexities that are introduced into selection studies whenever environmental factors may be involved. From the standpoint of the mother, the loss of a child from postnatal asphyxia is probably not so very different from having a stillborn child, or from the death of a child shortly after birth from erythroblastosis or haemorrhagic disease. It seems unlikely, therefore, that the effects on her reproductive behaviour will be strikingly different in the case of postnatal asphyxia. The true reason for the association with an elevated fertility might therefore be that his latter condition tends to occur preferentially in an unusually fertile subgroup of the population.

This view is supported in the case of postnatal asphyxia by evidence from the parental age and birth order studies discussed earlier. The risk was shown to be substantially elevated among children of very young mothers when contributions from birth order effects were excluded, and also among children of higher birth orders when contributions from maternal age effects were removed (Tables 1 and 3). This seemingly paradoxical relationship is easier to visualize when one considers that maternal

TABLE 7

Reproduction following deaths of infants from erythroblastosis (*born* 1953-55)

Birth order	Number of deaths	Mothers' ages	Subsequent births over 4 years * Exp.	Obs.
1	5	20 20 20 21 28	4·10	4
2	15	20 21 23 23 24 25 26 27 28 29 30 30 31 33 38	7·54	9
3	8	23 26 28 31 33 33 34 37	2·63	2
4	8	24 24 26 28 29 29 31 35	3·22	2
5	5	21 27 32 33 36	1·69	1
6–8	7	33 33 33 33 36 41 41	1·75	3
9–	4	30 30 33 34	1·14	1
Combined	52		22·07	22 †

* I.e., to end of 4-year period following the birth of the dead infant, but not beyond 1958.

† Includes 3 stillbirths and 2 infants dying before the end of the 4-year period.

Av. age = 1,514/52 = 29·12.

TABLE 8

Reproduction following the deaths of infants from haemorrhagic diseases (*born* 1953–55)

Birth order	Number of deaths	Mothers' ages	Subsequent births over 4 years Exp.	Obs.
1	5	18 20 20 26 27	4·44	4
2	7	18 22 25 26 28 28 29	4·63	5
3	6	27 28 30 34 38 40	1·86	2
4	4	25 28 29 39	1·29	1
5	4	25 34 36 36	1·44	2
6–8	3	26 35 37	0·37	0
9–	0	...	...	...
Combined	29	...	14·03	14*

* Includes 1 stillbirth and 1 infant dying before the end of the 4-year period.

Av. age = 834/29 = 28·76.

TABLE 9

Reproduction following deaths of infants from asphyxia (*born* 1953–55)

Birth order	Number of deaths	Mothers' ages	Subsequent births over 4 years Exp.	Obs.
1	27	16 18 18 18 19 20 20 21 21 21 22 22 22 23 23 23 23 24 24 26 28 29 32 32 32 39 40	22·76	28
2	20	18 20 21 22 23 23 24 24 26 26 27 27 28 29 30 33 35 35 38 41	12·71	22
3	20	18 21 22 22 22 23 23 23 25 26 27 27 27 27 29 30 30 31 36 43	13·22	25
4	15	20 23 24 26 26 27 27 27 28 28 29 32 33 34 35	8·90	17
5	8	22 22 24 25 26 27 28 31	5·18	10
6–8	7	21 26 30 31 34 38 38	3·19	9
9–	0	...	...	...
Combined	97	...	65·96	111*

* Includes 2 stillbirths; no liveborn infants died before the end of the 4-year period.

Av. age = 2,580/97 = 26·60.

age and birth order interact in such a way as to exaggerate the risk to a child of high birth order when it is born to a very young mother (as, for example, a third child to a nineteen-year-old girl), and also the risk to a child of low birth order when it is born to an older mother (as, for example, a first or second child to a woman of thirty-five or over). Stated in another way, it is risky for the child if its birth order is strikingly unusual in view of the mother's age.

All this implies that there are strong intra-uterine environmental influences operating on the foetus, and that a breakdown of the data by social factors might conceivably serve to indicate effects of the environments to which the mothers themselves are exposed. Thus, analyses in terms of social variables will eventually be needed if genetic conclusions are to be drawn from selection studies of hereditary conditions whenever these are influenced in their expressions by environmental factors. Ease of interpretation is not to be expected, but to follow the superficially simple alternative of studying only those traits that are determined wholly by genetic factors, and not at all by the environment, would exclude from consideration the bulk of the important hereditary and partially hereditary conditions of man.

Curiously, for one well-known genetic condition, namely mongolism, that is believed to be determined entirely by inherited factors, there is an elevated maternal fertility following the birth of an affected child. The 118 mothers of mongol children in the present study produced a total of forty-five subsequent offspring in the period under consideration, as against an expected thirty-five, an excess of 27 per cent (Table 10).

Just why this should be so is not clear. There is reasonably good ascertainment of cases of mongolism in British Columbia so that more nearly complete reporting of affected children from the lower, and more fertile, socio-economic subgroups is not a likely explanation. No one has seriously proposed that the risk of mongolism, for a given age of mother, is greater in these subgroups than for the rest of the population, but this is a possibility.

Such a suggestion should not be taken to imply that mongolian idiocy may be caused by an environmental effect upon the foetus in the absence of the gross chromosomal anomalies usually responsible. But effects of the external environment upon the mother could perhaps serve to increase the frequency of chromosomal non-disjunction in her reproductive cells, in much the same way that ageing does. Such an effect would tend to escape detection because of a lower average age of motherhood in lower socio-economic groups, and its demonstration would depend upon comparisons being made of the age specific risks in different population subgroups.

TABLE 10

Reproduction following the births of mongoloid infants (born 1953–55)

Birth order	Number of cases	Mothers' ages	Subsequent births over 4 years: Exp.	Obs.
1	21	17 17 18 19 19 21 21 22 22 23 24 25 25 27 29 34 39 39 41 41 42	15·08	20
2	18	22 25 27 28 28 30 33 36 37 38 38 39 40 40 41 41 41 42	5·04	4
3	35	19 26 28 28 30 31 31 32 32 33 33 33 34 34 34 34 34 35 36 36 37 38 38 39 40 40 40 41 41 41 41 41 42 43 47	8·14	10
4	23	23 25 30 30 31 31 34 34 35 36 37 37 37 38 38 39 40 40 41 42 43 45 48	4·81	7
5	7	29 33 34 39 40 40 46	1·08	0
6–8	7	35 37 38 39 40 43 46	0·77	0
9–	7	39 40 41 41 43 43 46	0·46	4
Combined	118	...	35·38	45*

* Includes 2 infants dying before the end of the 4-year period; no stillbirths occurred in this period.

Current findings from studies of selection thus serve to emphasize that increased availability of empirical information on fertility differences brings with it unexpected problems of interpretation and that, for the solution of these, further knowledge of the social characteristics of the individuals is required. Fortunately, our religious, ethnic and economic particulars are fairly well documented, even if only the vital records are considered, and an

TABLE 11

Under- and over-fertility following infant mortality and morbidity (measured over a 4-year period)

Condition	Number of cases*	Subsequent births			Surviving to end of 4 years		
		Obs.	Exp.	Ratio	Obs.	Exp.	Ratio
Stillbirths †	269	136	146·0	0·93	...	...	...
Erythroblastosis (deaths)	52	22	22·1	0·99	17	21·3	0·80
Haemorrhagic disease (deaths)	29	14	14·0	1·00	12	13·5	0·88
Mongolism	118	45	35·4	1·27	43	33·6	1·28
Asphyxia (deaths)	97	111	66·0	1·68	109	63·0	1·73

* Stillbirths in 1953, and births in 1953–55 of children with other conditions.
† Previous data (Newcombe [11]) cases are limited to birth orders 1 to 4 inclusive.

immediate task is the fuller utilization of the information that is already potentially available, on a scale appropriate to permit multiple breakdowns of the data.

As a minor refinement in the present study, the effective fertilities of the mothers in terms of the numbers of offspring surviving to the end of the period under consideration have been derived for the four conditions, erythroblastosis, haemorrhagic disease, postnatal asphyxia and mongolism (Table 11). When this is done, a slight reduction of fertility is observed in the case of the first two conditions, and a slight increase in the case of the other two.

Further attempts to interpret the observed effects would be unprofitable at this stage when we have only just started to describe them. The study so far has considered only five conditions, only 500 to 600 affected infants, and the reproductive performances of the mothers over a period of only four years maximum. The methods are applicable equally to conditions of other kinds and to much larger numbers of families followed over much longer time spans.

Disease Conditions for which Studies of Mutation and Selection are Profitable

The empirical data yielded by studies of mutation and selection are, of course, most simply interpreted where the traits under investigation are readily recognizable, are determined by single gene differences and are always expressed when the causal gene is present. For this reason, such studies have in the past tended to make use of a few quite rare genetic traits that come reasonably close to meeting the ideal specifications. Achondroplastic dwarfism and Huntington's chorea are examples in point. While appropriate for establishing genetic principles, this sort of approach concentrates, as it were, on a small and unrepresentative part of the total genetic variability affecting man's well-being.

At the other end of the spectrum, attempts to estimate the likely total magnitudes of the consequences of mutation and of selection have sometimes been based on gross mortality data, not subdivided at all by disease cause.[2, 5]

In between the two extremes there is a wide range of possible lines of study awaiting exploration, information on which is abundantly available even if a little tricky to get at. While routine records such as those of deaths and of hospital discharges are correctly regarded as inefficient sources of ascertainments of cases of the rarer genetic diseases, they do contain diagnoses of other sorts of conditions that are appropriate for studies either of mutation or of selection. Very large numbers of blood groupings, for example, are recorded annually in hospitals and as a part of the running of blood banks and transfusion services. Cases of important diseases that are partially genetic in origin and inherited in an irregular fashion are likewise reasonably well documented, although not necessarily in any single set of files; diabetes, idiopathic epilepsy and certain of the congenital malformations being notable examples. A number of the metrical, or quantitative, traits of man, of kinds that are suitable for studies of selection pressures, are likewise widely recorded; examples include birth weight which is stated on virtually all birth registrations, intelligence which is

measured routinely in many school systems, and arterial pressure which is commonly taken in the course of medical examinations.

The possible applications of linkages with family pedigree information, such as may be obtained from the vital registrations, varies somewhat with the nature of the trait under study. In the case of the blood groups an important aim is the detection of correlations with fertility, mortality and morbidity, which will help to account for the continued prevalence of a diversity of gene alleles. In the case of birth weight variations there is a rather special unsolved problem, in that the modal value differs unexpectedly from the selective optimum as judged from the abilities of the infants to survive the first year of life.[3] There are also special problems associated with the study of selection as relating to intelligence and school performance, in that most investigations have in the past been concerned mainly with a single generation of test scores. The refinements that might derive from use on a substantial scale of two or more generations of scores within the same pedigrees, to help disentangle a transmissible component in mental ability from the more obviously environmental components for the purposes of selection studies, has hardly been explored at all even in theoretical discussions of the degree of precision that might eventually be possible.

For the diseases that display irregular inheritance another sort of problem arises, that bears on the manner in which tests for selective pressures may be designed, and that can in part be investigated by the use of pedigree information. Because of the nature of these conditions, many of the carriers of the causal genes may not be affected, and still others may have diseases that do not especially resemble the usual expressions of the genes. A few aetiological associations between outwardly diverse traits of an hereditary or partially hereditary sort have already been established, as in the case of spina bifida and anencephalus, and where these exist it is desirable to study the effects of selection on all of the known expressions of the underlying causal factors.

That such aetiological groupings may perhaps be more common than has been recognized in the past is suggested by a recent study of the risks of death and disease among the brothers

TABLE 12

Handicaps, stillbirths and deaths among sibs of children with strabismus

(Based on files relating to children born in British Columbia during 1953–58 and registered as stillborn, and registered as handicapped before the end of 1960 or as dying before the end of 1959)

Conditions	Cases		Ratio obs./exp.	Excess obs. minus exp.	
	Obs.	Exp.			
Among 207 liveborn plus stillborn sibs, born after the first child with strabismus					
Strabismus	13	0·27	47·62	12·73	
Other registered handicaps	7	4·32	1·62	2·68	
Stillbirths	9	2·31	3·90	6·69	15·20
Deaths	12	6·18	1·94	5·83	
Combined	41	13·08	3·14	27·93	
Among a total of 310 liveborn plus stillborn sibs, born before and after the first child with strabismus					
Other registered handicaps	11	6·48	1·70	4·52	
Stillbirths	10	3·45	2·90	6·55	
Deaths	17	9·25	1·84	7·75	
Combined	38	19·18	1·98	18·82	

Note: The expected numbers are based on the numbers of registered cases of strabismus (286 cases representing 273 families), other handicaps (4,509), stillbirths (2,403) and deaths to the end of 1959 (6,437) out of the total birth population, i.e. live plus still, for the 6-year period 1953–58 (215,795).

and sisters of children with strabismus. This irregularly inherited trait is particularly well reported in British Columbia, so that the cases appearing in the Register of Handicapped Children and Adults represent a substantial fraction of the total. In the 207 affected families which were studied using the linked files of births, deaths and handicaps, the risk of strabismus among the later born brothers and sisters was between forty and fifty times that for the birth population as a whole (i.e. 6·3 per cent as compared with 0·13 per cent; see Table 12). This, however, was not the only risk to which the families were exposed.

An unexpected finding of the study was that of an approximate doubling of the risks of stillbirth, death and of handicapping conditions of other kinds. In fact, the excess of these other sorts of casualty, taken together, equalled in number the strabismus cases, each being represented by about six affected children per 100 brothers and sisters. Thus strabismus would appear to be aetiologically related in some way with a wide range of conditions, and about half of the special risk to the later born siblings of a child who has strabismus come from the aetiologically related conditions that may not even remotely resemble strabismus (see Table 13).

How then do we define in this instance the genetic entity which we wish to study? If selection is shown to operate in a certain way when individuals or families with strabismus are investigated, might it not perhaps operate in a different way when the genes responsible, or some part of them, are expressed differently? The interpretation of empirical data on selection pressures is in such circumstances complicated by the possibility that there may be still further expressions of the genes that have not so far been identified; but this is preferable to the possible over-simplifications that might arise from recognition of only a single kind of expression of the genes.

The exploration of aetiological associations has in the present instance been aided by two circumstances: (*a*) the use of systematically collected files of all live births, stillbirths, child handicaps and child deaths in a large defined population, regardless of any preconceived notions concerning the genetic or non-genetic origins of the various events of morbidity and mortality; and, (*b*) the integration of these files into the form of family histories of vital and health events using rapid modern methods of data processing. Under no other circumstances could the relationships with strabismus have been demonstrated with the same degree of precision, or in the relatively short period of time (about two days) which the study took. Current work is being directed to eliminating some remaining manual steps in such operations so that a large number of conditions can be treated similarly in a single study using larger files of records.

TABLE 13

Causes of stillbirth, handicap and death among 310 *sibs of children with strabismus (excluding cases of strabismus)*

Code	Cause	Cases
Stillbirths		
Y35·3	Ill-defined causes in the mother	1
Y36·1	Placenta praevia	2
Y36·2	Prematurity	2
Y36·6	Abnormality of the placenta	1
Y38·1	Hydrocephalus	1
Y39·6	Unspecified	2
Y38·6, Y38·2	Monster, spins bifida	1
Handicaps		
004·0	Tuberculosis	1
326·2	Behaviour disorder	1
351·0	Cerebral palsy	1
389·4	Retrolental fibroplasia	1
748·0	Clubfoot	1
754·4	Fibroelastosis cordis	1
757·2	C.m. of external genitalia *	1
325·4, 748·0	Mongolism, clubfoot	1
325·5, 351·0	Mental deficiency, cerebral palsy	2
759·3, 758·6	Multiple malformations, including skeletal	1
Deaths		
057·0	Meningococcal infections	1
330·0	Subarachnoid haemorrhage	1
490·0	Lobar pneumonia	1
493·0	Pneumonia	1
754·2	Interventricular septal defect	2
754·4	Fibroelastosis cordis	1
762·5	Postnatal asphyxia, with immaturity	2
763·0	Pneumonia of the newborn	1
771·5	Haemorrhagic disease, with immaturity	1
776·0	Immaturity	4
910·0	Accidental injury	1
924·0	Accidental suffocation	1

* C.m. = congenital malformation.

Costs and Administrative Applications

I will not at this stage discuss in any detail the methods used for this kind of work, since these have already been fully described elsewhere.[4, 5, 9, 10, 11] The present status of the undertaking has also been reported earlier.[8]

However, since descriptions of this kind of work may give it the appearance of being excessively elaborate and costly in relation to the likely yield of genetic information, some clarification may be in order. The record linkage approach makes use only of information that is already being collected and recorded. In our own case, it has also employed mechanically readable punchcards that are already being prepared routinely from these original records for the extraction of annual statistics, and for the compilation of alphabetic name indexes of the kinds used for searching out birth registrations and such. The only innovation in relation to these existing procedures has been a minor modification of the vital statistics name index cards to include additional family identification such as the mother's maiden surname and the parental initials, ages and provinces of birth. The cost involved is slight, and the modified index cards, suitable for purposes of family linkage, are now used routinely for all events of marriage, live birth, stillbirth and death over the whole of Canada.

The most demanding part of such an undertaking is the development of increasingly versatile and rapid computer methods for inter-filing and linking the successive annual crops of records into family packets, and for extracting the resulting pedigree information in the required form after this has been done. Using the most recent programmes developed for a fast modern computer, the cost of the actual merging and linking of the marriage, birth, death and handicap records, when suitably arranged in magnetic tape form, is very much less than that of preparing the punchcard records in the first place, being only about a quarter of a cent per updating record; and the current speeds of the updating and linking operation are in the vicinity of 1,000 to 2,000 incoming records per minute, so that the 40,000 annual birth records for British Columbia can be incorporated into the master family file in less than an hour.

What may not be generally recognized is the extent to which methods for the automatic linkage of records may serve existing administrative needs, relating to the uses of these same records. One example will serve to illustrate the point. The bringing together of information from all birth registrations into family groupings is, in fact, already being carried out over the whole of Canada by laborious manual methods as a necessary part of the operation of a government scheme of family allowances. To ensure that false claims will be detected, information from the mother's application form is in each instance compared visually with alphabetic lists of the names of newborn infants derived from the birth index cards, and with information from the existing family files wherever the family is already "in pay". These verifications of the mothers' statements account for 20 per cent of all family allowance file upkeep operations, which together cost in the vicinity of $1.25 per family per year. Substantial financial savings are thus potentially possible, using a rapid automatic record linkage procedure which could in addition yield family groupings of records for population research.

There are numerous other examples of uses to which the linkage technology might be put, and if family groupings of routine records are ever to be employed extensively as a source of information on the genetics of human populations, it will be because cognizance has been taken of these administrative needs in setting up stable, continuing systems which will permit scientific information to be derived as a by-product.

In the long run, it is doubtful whether precision in the study of selective forces, and perhaps in the study of mutation as well, is likely to be achieved in any other way, in view of the inaccuracies and the limited quantity of family pedigree information that can be derived from *ad hoc* surveys of conventional design based on personal interviews and questionnaires. Precision in the study of mutation and selection will undoubtedly be needed if we are to understand, and eventually control as Galton

suggested, the influences that determine our genetic future.

Summary

Studies of mutation and selection, as influencing the genetically simple traits of man, have been popular in the past but they tell little about the overall magnitudes of the effects of these two forces throughout the genome as a whole. Similar studies would be desirable, relating to the collectively more important characteristics of man that are less simply inherited, and that are determined by combinations of environmental and hereditary factors. But such work is beset with difficulties arising out of the numerous environmental inhomogeneities characteristic of human populations. To be interpretable, data from these more difficult sorts of investigations must be broken down by a multiplicity of social variables; and correspondingly larger amounts of data are needed to make possible the required multivariate approach.

An abundance of relevant information about people is in fact recorded, which we have only begun to tap, describing not only their biological characteristics and family relationships, but their social circumstances as well. Application of this body of knowledge depends largely upon bringing together, or "linking", the separately recorded facts about the same individuals and families. Such "record linkages" will have many uses if the process of linking can be made cheap enough and rapid enough. In the long run, understanding of the magnitudes of the effects of recurrent mutations, and of various selective forces, on the quality of the human gene pool may depend upon the use of such methodological developments to take maximum advantage of information we already possess.

REFERENCES

1. Cavalli-Sforza, L. L. 1962. Demographic attacks on genetic problems: some possibilities and results. In *The Use of Vital and Health Statistics for Genetic and Radiation Studies.* Pp. 221–30. United Nations, New York.
2. Crow, J. F. 1962. Selection. In *Methodology in Human Genetics.* Pp. 53–75. Ed. W. J. Burdette. Holden-Day Inc., San Francisco.
3. Karn, M. N., and Penrose, L. S. 1952. Birthweight and gestation time in relation to maternal age, parity and infant survival. *Ann. Eugen. Lond.* **16**, 147.
4. Kennedy, J. M. 1961. *Linkage of Birth and Marriage Records Using a Digital Computer.* Document no. A.E.C.L.-1258, Atomic Energy of Canada Limited, Chalk River, Ontario.
5. Kennedy, J. M. 1962. The use of a digital computer for record linkage. In *The Use of Vital and Health Statistics for Genetic and Radiation Studies.* Pp. 155–60. United Nations, New York.
6. Morton, N. E., Crow, J. F., and Muller, H. J. 1956. An estimate of the mutational damage in man from data on consanguineous marriage. *Proc. nat. Acad. Sci., Wash.* **42**, 855.
7. Newcombe, H. B. 1964. Screening for effects of maternal age and birth order in a register of handicapped children. *Ann. hum. Genet. Lond.* **27**, 367.
8. Newcombe, H. B. 1964. Pedigrees for population studies. A progress report. *Cold Spr. Harb. Symp. quant. Biol.* **29**, 21.
9. Newcombe, H. B., and Kennedy, J. M. 1962. Record linkage: Making maximum use of the discriminating power of identifying information. *Communications of the Association for Computing Machinery*, **5**, 563.
10. Newcombe, H. B., Kennedy, J. M., Axford, S. J., and James, A. P. 1959. Automatic linkage of vital records. *Science*, **130**, 954.
11. Newcombe, H. B., and Rhynas, P. O. W. 1962. Child spacing following stillbirth and infant death. *Eugen. Quart.* **9**, 25-35.
12. Newcombe, H. B., and Tavendale, O. G. 1964. Maternal age and birth order correlations. Problems of distinguishing mutational from environmental components. *Mutation Research*, **1**, 446.
13. Newcombe, H. B., and Tavendale, O. G. 1965. Effects of father's age on the risk of child handicap and death. *Amer. J. hum. Genet.*, **17**, 163.
14. Penrose, L. S. 1955. Parental age and mutation. *Lancet*, ii, 212.
15. *Recommendations of the International Commission on Radiological Protection.* 1959. Pergamon Press, New York.
16. Sonneborn, T. M. 1956. Paternal age and stillbirth rate in man. *Records of the Genetics Society of America*, Number 25, p. 661.